Big data - modelling of midges in Europa using machine learning techniques and satellite imagery

Cuellar, Ana Carolina; Kjær, Lene Jung; Skovgaard, Henrik; Nielsen, Søren Archim; Stockmarr, Anders; Anderson, G.; Lindström, Anders; Chirico, J.; Lilja, T.; Lühken, R.

Total number of authors:
36

Published in:
NKVet Symposium 2017 - abstract book

Publication date:
2017

Document Version
Peer reviewed version

BIG DATA - MODELLING OF MIDGES IN EUROPA USING MACHINE LEARNING TECHNIQUES AND SATELLITE IMAGERY

Cuellar A.C.[1], Kjær, L.J[1], Skovgaard H. [2], Nielsen S.A.[3], Stockmarr A.[4], Anderson G.[5], Lindstrom A. [5], Chirico J. [5], Lilja T.[5], Lühken R. [6], Steinke S. [6], Kiel E.[6], Larska M.[7], Hamnes S.I.[8], Sviland S. [8], Hopp P. [8], Brugger K.[9], Rubel F.[9], Balenghien T.[10], Garros C. [10], Rakotoarivony I.[10], Allène X. [10], Lhoir J. [10], Delécolle J.C. [11], Mathieu B. [11], Delécolle D. [11], Setier-Rio M.L. [12], Venail R. [12], Scheid B. [12], Miranda Chueca M.A. [13], Barcelo Segui C. [13], Lucientes J.[14], Estrada R.[14], Wesley Tack[15], Mathis A.[16], and Bødker R.[1]

[1]Section for Diagnostics and Scientific Advice, Technical University of Denmark (DTU), National Veterinary Institute, Copenhagen, Denmark
[2] Department of Agroecology - Entomology and Plant Pathology, Aarhus University, Aarhus, Denmark
[3] Institute for Science and Environment, Roskilde University, Roskilde, Denmark
[4]Department of Applied Mathematics and Computer Science, Technical University of Denmark (DTU), Copenhagen, Denmark
[5]National Veterinary Institute (SVA), Uppsala, Sweden
[6]Bernhard Nocht Institute for Tropical Medicine, WHO Collaborating Centre for Arbovirus and Haemorrhagic Fever Reference and Research National Reference Centre for Tropical Infectious Diseases, Hamburg, Germany
[7]Department of Virology, National Veterinary Research Institute, Pulawy, Poland
[8]Norwegian Veterinary Institute, Oslo, Norway
[9] Institute for Veterinary Public Health, Vienna, Austria
[10]Agricultural research for development (CIRAD), Montpellier, France
[11]Institute of parasitology and tropical diseases of Strasbourg, Université de Strasbourg, Strasbourg, France
[12] EID Méditerranée, Montpellier, France
[13] Laboratory of Zoology, University of the Balearic Islands, Spain
[14] Department of Animal Pathology, University of Zaragoza, Spain
[15]Avia- GIS, Zoersel, Belgium
[16] Institute of Parasitology, University of Zürich, Zürich, Switzerland

Corresponding author: anacu@vet.dtu.dk

Biting midges (Diptera, Ceratopogonidae) of the genus *Culicoides* are important vectors of pathogens causing diseases in free living and production animals and can lead to large economic losses in many European countries. In Europe, *Culicoides imicola* and the *Obsoletus* group are considered to be the main vectors of bluetongue virus that mostly affects ruminants such as cattle and sheep. Spatio-temporal modelling of vector distribution and abundance allows us to identify high risk areas for virus transmission and can aid in applying effective surveillance and control measures.

We used presence-absence and monthly abundance data of *Culicoides* from 1005 sites across 9 countries (Spain, France, Denmark, Poland, Switzerland, Austria, Poland, Sweden, Norway) collected between the years 2007 and 2013. The dataset included information on the vector species abundance (number of specimens caught per night), GPS coordinates of each trap, start and end dates of trapping. We used 120 environmental predictor variables together with Random Forest machine learning algorithms to predict the overall species distribution (probability of occurrence) and monthly abundance in Europe. We generated maps for every month of the year, to visualize the abundance of *C. imicola* and *Obsoletus* group in Europe as well as distribution maps showing the probability of occurrence.

We were able to create predictive maps of both *Culicoides* sp. occurrence and abundance using Random Forest models, and although the variance was large, the predicted abundance values for each site had a positive correlation with the observed abundance. We found relatively large spatial variations in probability of occurrence and abundance for both *C. imicola* and the *Obsoletus* group. For *C. imicola* probability of occurrence and abundance was higher in southern Spain, where as the *Obsoletus* group had higher probability of occurrence and abundance in central and northern Europe such as France and Germany. Temporal variation was also observed with higher abundance occurring during summer months and low or no abundance during winter months for both *C. imicula* and the *Obsoletus* group, although abundance was generally higher for a longer period of time for *C. imicula* than for the *Obsoletus* group.

Using machine learning techniques, we were able to model the spatial distribution in Europe for *C. imicola* and the *Obsoletus* group in terms of abundance and suitability (probability of occurrence). Our maps corresponded well with the previously reported distribution for *C. imicola* and the *Obsoletus* group. The observed seasonal variation was also consistent with reported population dynamics for *Culicoides*, as it depends on environmental factors such as temperature and rainfall. Longer seasonal abundance for *C. imicula* compared to the Obsoletus group can be explained by the species distribution, as *C. imicula* is limited to the southern parts of Europe where the warm season lasts longer, whereas the *Obsoletus* group is found further north. The outputs obtained here will be used as input for epidemiological models and can be helpful for determining high risk areas for disease transmission.